Report No. UT-03.24

HIGHWAY ACCIDENT EXPECTED VALUE ANALYSIS

By: Peter Martin, Ph.D.
Yali Cheng
Aleksandar Stevanovic

Civil & Environmental
Engineering Department
University of Utah
Salt Lake City, Utah

Utah Department of Transportation
Research Division

September 2003

UDOT RESEARCH & DEVELOPMENT REPORT ABSTRACT

1. Report No. UT- 03.24	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Highway Accident Expected Value Analysis	**5. Report Date** September 2003 **6. Performing Organization Code**	
7. Author(s) Peter Martin, Ph.D Yali Chen Aleksander Stevanovic	**9. Performing Organization Report No.**	
9. Performing Organization Name and Address University of Utah 1471 Federal Way Salt Lake City, Utah 84102	**10. Work Unit No.** **11. Contract No.** 039151	
12. Sponsoring Agency Name and Address Stan Burns, P.E. Utah Department of Transportation 4501 South 2700 West Salt Lake City, UT 84114-8410	**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code**	
15. Supplementary Notes		

16. Abstract

The Utah Department of Transportation (UDOT) must have accurate accident statistics for different highway classes to improve the safety of the Utah highway system. Although UDOT keeps a complete accident database, it does not reveal commonalities and trends behind accidents. In addition, Utah traffic facilities and services change over the years, emphasizing the need to evaluate the Utah Highway Networks' Functional Classification System.

This study measures the safety of different Utah highway functional classes (FCs) according to traffic patterns, traffic volume, geometrics, and travel speed. It presents a process to analyze accident data for different FCs. It also produces the expected accident range in order to identify locations with high accident rates.

17. Key Words Accident Expected Accident Rate, Accident Rate		**18. Distribution Statement** Available: UDOT Research Division Box 148410 Salt Lake City, UT 84114-8410 www.udot.utah.gov	
19. Security Classification (of this report) N/A	**20. Security Classification (of this page)** N/A	**21. No. of Pages** **19**	**22. Price** N/A

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Acronyms

AADT – Average Annual Daily Traffic

CARS – Centralized Accident Records System

CSM – Classic Statistical Method

EV – Expected Value

EVA – Expected Value Analysis

FC – Functional Class

UDOT – Utah Department of Transportation

SQL – Structured Query Language

EXECUTIVE SUMMARY

The Utah Department of Transportation (UDOT) must have accurate accident statistics for different highway classes to improve the safety of the Utah highway system. Although UDOT keeps a complete accident database, it does not reveal commonalities and trends behind accidents. In addition, Utah traffic facilities and services change over the years, emphasizing the need to evaluate the Utah Highway Networks' Functional Classification System. This study measures the safety of different Utah highway functional classes (FCs) according to traffic patterns, traffic volume, geometrics, and travel speed. It presents a process to analyze accident data for different FCs. It also produces the expected accident range in order to identify locations with high accident rates.

1. INTRODUCTION

UDOT keeps a database of traffic accidents that occurred between 1997 and 2001. However, it does not currently provide an accurate expected accident range. The expected accident range shows whether an accident rate is acceptable in a certain location or whether the rate significantly exceeds the expected range.

Common statistical methods to analyze traffic accident patterns are spot maps, accident frequency, accident rate, the rate quality control method, the Bayesian method, the classic statistical method, and Expected Value Analysis (EVA). Each of these procedures was developed to identify hazardous locations.

Accident locations are marked on spot maps to determine accident-prone areas. They are especially useful in tracking pedestrian accidents and parked car accidents on urban streets (2).

Accident frequency is a simple method that identifies locations with high accident rates by ranking the frequency at which accidents occur in the same location. Like the accident frequency method, the accident rates method ranks locations by accident rate. Accident rate represents the number of accidents in relation to the volume of traffic entering a section or intersection (2).

The rate quality control method is a scientific approach to traffic accident analysis. Unlike the classic statistical method and EVA, it assumes that the traffic accident number at a set of locations follows a Poisson distribution. It analyzes the rate, not the frequency, of accidents and compares the accident rate of a particular location to the mean rate at similar locations (2).

The Classic Statistical Method (CSM) and EVA are based on normal distribution, or the frequency at which things occur in the world. Normal distribution has applications in traffic safety issues, as many variables used to describe traffic systems are normal or follow some form that can be derived from normal distribution. Normal distribution is a simple statistical tool because it is used widely in traffic applications.

CSM and EVA identify locations with high traffic accident rates by comparing accident rates and the mean accident rate for a given location. Using a one-sided confidence interval, CSM emphasizes the accident rates that are significantly higher than the mean accident rate. EVA focuses on the abnormalities in location accident rate and compares the accident rate from a certain location with the mean accident rate using a two-sided confidence interval.

EVA compares sites with similar patterns of traffic control, geometrics, speed, volume, and density. It locates areas with similar geometric and traffic characteristics and determines their mean number of specific accident types. These averages predict accident rates at specific location types. Table 1 provides data collected at ten segments of roadway, with similar geometric characteristics and traffic conditions. The table shows how EVA derives a mean number of accidents for such sites.

Table 1: EVA Example

Control Site	Number of Accidents
1	31
2	33
3	32
4	28
5	33
6	32
7	28
8	30
9	26
10	30
Average	30.3
Standard deviation	2.36

The objective is to determine whether the study site is hazardous for a 95% confidence level (Z = 1.96).

The expected range is equal to: $E(X) = \overline{X} \pm ZS$ = 30.3 ± 2.36 * 1.96 = 34.92 < 35.

The study site is dangerous at this level.

2. OBJECTIVES AND TASKS

UDOT decision makers need information about accident statistics for different highway classes to improve highway safety. UDOT keeps a complete accident database. However, the database does not reveal the commonalities and trends behind accident data. This project develops a methodology for using expected accident range to identify high-accident locations. It uses EVA to analyze accident patterns in different highway FCs, or highways with specific services.

The project objectives are as follows:

1. Determine whether the combinations of highway FC can be used to adequately represent the Utah Highway System with respect to accident analysis.
2. Find the expected accident range for each highway FC.
3. Recommend the process used to calculate the expected accident ranges, deliver it to users, and maintain the range table over time.

Accident rate varies according to changes in geometrics, traffic volume, traffic speed, traffic control, and traffic density. Range tables must be modified to reflect these changes. This project provides a process for renewing range tables.

This project completes the following tasks:

1. Uses UDOT's Centralized Accident Records System (CARS) database to identify whether the accident rates for each highway FC follow a normal distribution.
2. Develops methods to handle irregularities in distribution of accidents per highway FC.
3. Recommends modifications in UDOT's highway FC system based on accident data analyses to better represent Utah highways.
4. Employs statistical methods to segment the population in each highway FC into ranges representing low, expected, and high accident rate levels.
5. Provides a methodology used for future UDOT analysis and calculation of expected accident ranges.
6. Recommends policies and procedures for UDOT to calculate expected accident range.

3. LITERATURE REVIEW

The literature evaluates methodologies and computer applications for calculating accident statistics. South Dakota, Virginia, Michigan, and Texas have produced EVA tables. Many of these deal with similar intersection accident rates.

3.1 Methods

(1) Northwestern University Traffic Institute Traffic Engineering Analysis
This analysis presents several methods used to identify high accident locations: Number of Accident Method, Accident Rate Method, Number-Rate Method, and Rate Quality Control Method. Neither Accident Rate Analysis nor Number of Accidents is completely accurate in identifying hazardous locations. However, when the two methods are combined, deficiencies are minimized or eliminated.

(2) Statistical Quality Control Techniques
The Rate Quality Control Method was developed in 1956. It analyzes highway accident data using statistical quality control techniques (4). It determines whether accident rates are normal or abnormal in relation to a predetermined mean accident rate for similar locations using statistical tests (10).

(3) Bayesian Method

a. Estimating Safety by the Empirical Bayes Method: Tutorial

The Empirical Bayes method increases the precision of accident estimates and corrects the regression-to-mean bias (5).

b. Bayesian Identification of Hazardous Locations, Julia L. Higle and James M. Witkowski

Analyzing hazardous accident locations using accident rate data appears to be a sound procedure (6). The Higle-Witkowski Bayesian model offers many advantages over standard methods. However, the most accurate analytical methods are not yet available.

3.2 Evaluation of the Conventional Accident Rate Measure

Accident rates are commonly used to analyze hazardous roadway locations. However, rates do not accurately reflect the true degree of hazard because of aggregation effects (7).

3.3 Automated Expected Value Analysis for Accident Rate

T. Chira Chavala and King K. Mak developed an algorithm to identify factors that cause accident overrepresentation at a site. Engineers can use the output of the algorithm to respond to problems at a site and to develop a traffic safety improvement plan for the area's highway network (8).

King K. Mak, T. Chira-Chavala, and Barbara A. Hilger developed MAAP, a microcomputer program. This program is currently being field tested at a small number of sites in Fort Worth, Houston, and San Antonio, Texas. The program offers the following options.

(1) Years of accident data
(2) Accident selection (sub-setting) criteria, such as county location, highway type, accident type, and accident severity

3.4 Expected Value Range Tables From Other States

South Dakota developed EVA tables from accident data collected at random intersections throughout the state. Since the EVA tables are reliable, they will assist the South Dakota Department of Transportation (SDDOT) in identifying abnormal accident patterns at specific intersections. If a certain type of intersection seems to have a high number of a particular type of accident, the corresponding table can help confirm whether this is in fact true. Safety precautions can then be taken to reduce the number of accidents (11).

Oregon Department of Transportation (ODOT) keeps a five-year record of crash rates. Each state highway is divided into urban and rural sections. The sections are then subdivided for analysis. The record shows the start-mile point, segment length, number of crashes, average daily traffic, current crash rate, and the crash rate for the previous four years in each sub-section (12).

4. HIGHWAY FUNCTIONAL CLASSIFICATION (FC) EVALUATION

Highway FC groups highways according to the traffic services they provide. This classification aids in administrating highway systems, building and maintaining highways according to consistent design standards, and evaluating highway imperfections.

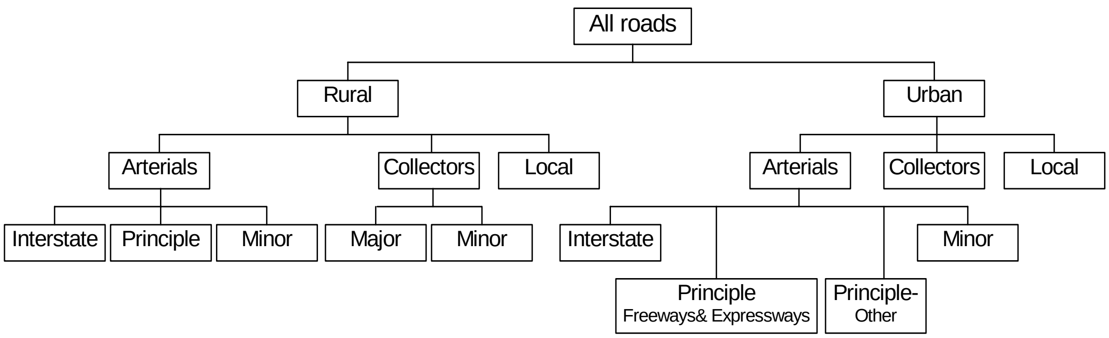

Figure 1: Utah Highway Functional Classification System

Figure 1 shows the hierarchy of the Utah Highway Functional Classification System. Utah highways are divided into three major classes: arterials, collectors, and local roads and streets. These three classes are further divided into urban and rural subclasses (13).

4.1 Rural Highway Functional System

Principal arterials facilitate corridor movements, including all highway trips between urbanized areas and a high percentage of trips between small urban areas. Generally, the highest traffic volume corridors, the longest continuous road trips, and the highest proportion of vehicle miles traveled take place on principal arterials. Principal arterials provide an integrated network of continuous routes that branch off into rural interstates and other arterials.

Minor arterials form an integrated network that connects cities, large towns and other traffic generators. This system provides service, at relatively high travel speeds, to corridors with trip lengths and travel density greater than those served by rural collectors.

Major collectors carry traffic to and from county seats, large towns, and county destinations such as consolidated schools, parks, or important mining and agricultural areas not served by an arterial. Minor collectors are spaced at intervals to collect traffic from local roads.

They link rural hinterland with local traffic generators, such as small communities. Local road systems provide travel access between adjacent areas. The system consists of all rural roads not included in the other system.

4.2 Urban Highway Functional System

Principal arterials serve major activity centers, highest volume corridors, and longest trip demands. They also connect with major rural corridors to accommodate entrance and exit of urban areas. This system accommodates a large portion of urban travel over a relatively small mileage zone. The system is separated into three subsystems:

1) Interstates consisting of principal arterials designated as part of the Interstate system, with fully controlled access and grade-separated interchanges
2) Freeways/expressways consisting of non-Interstate principal arterials with controlled access in addition to at-grade intersections
3) Other principal arterials without controlled access

Minor arterials connect to and augment urban primary arterials. They provide urban connections for rural collectors and provide more land access than principal arterials, without penetrating neighborhoods. They usually serve trips of moderate length. Urban collectors draw in traffic from local streets in residential areas or in CBDs and convey it to the arterial system. They penetrate residential areas and serve both land access and traffic circulation in residential and commercial/industrial areas. Local street systems include all streets within the urban area that are not included in other systems. These streets provide direct access to adjacent land and to larger systems. However, they do not offer through-traffic movement.

Highway FCs must be continuously evaluated due to frequent changes in highway networks. Traffic volume, highway reconstruction, physical distance, and travel speed define different FCs as well as determine highway safety. In this study, safety acts as an aggregate set to measure the accuracy of current Utah Highway FC.

5. EVA METHODS

5.1 Expected Value Analysis

This study uses EVA to produce an expected value range table for each highway FC.
Normal distribution of the events is generally used to measure traffic safety because it is simple, applicable, and useful. EVA assumes that accident rate frequency distribution follows normal distribution. The EV can be obtained from the following formula:

Formula 1

$$E(X) = \overline{X} \pm ZS \qquad (1)$$

$$\overline{X} = \frac{\sum X}{n} \qquad (2)$$

$$S = \frac{\sqrt{\sum (X - \overline{X})^2}}{n-1} \qquad (3)$$

Where:

$E(X)$ = Expected range of each highway FC

X = Accident rate of each segment belonging to similar highway FC

$\overline{X}$ = Average accident rate for the similar highway FC

Z = A probability constant corresponding to the required confidence level (In this study,
Z = 1.96)

S = Estimated standard deviation of the accident rate of each highway FC

n = Number of segments of each highway FC

5.2 Accident Analysis Methods

X represents either accident rate or number of accidents, depending on the available data. The following two methods can be used to find accident range:

1. The Accident Number Method is a simple and direct way to find accident range. Mean accident number and standard deviation can be used to calculate an abnormal upper limit rate on a certain level of confidence. Since this method does not consider the length of a highway segment or the number of vehicles using it, the results may not be representative of real world conditions (10).

2. The Accident Rate Method may be more reliable than the Accident Number Method because it considers traffic volume. This method accounts for the fact that one highway segment may have more accidents because it is used more than other segments. It uses the number of accidents and exposure data, such as traffic volume and the length of highway segment, to determine accident rate. Because accidents are rare events, accident rates are often small decimal fractions. In order to avoid

working with such small numbers, accident rates are multiplied by 1 million VMT and expressed as a rate per 1 million vehicles. The formula used is (4):

$$RMVM = \frac{A * 1{,}000{,}000}{365 * ADT * L}$$

Where:

RMVM = Number of accidents per million vehicle miles of travel
A = Number of total accidents or number of accidents during a study period
ADT = Average Daily Traffic
L = Length of road segment

After comparing the Accident Number Method and the Accident Rate Method, the rate method was found to be more effective for the purposes of this study.

6. DATA PRE-PROCESSING

6.1 Data Source

UDOT provides the CARS database, which catalogs detailed accident information from the years 1997 to 2001. The database consists of four tables. Table one is a road table that includes route numbers, begin-mile points, end-mile points, mile length, and Average Annual Daily Traffic (AADT). Table two is an accident table that records accident data, accident location, and accident type. Tables three and four describe the vehicles and people involved in specific accidents. CARS data produces an EV accident rate using data transformation, connection, and split.

6.2 Data Transformation

In this process, CARS data is loaded into Microsoft Access, the database management software for this project. Microsoft Access allows text files to be easily transformed into data files. The accident file includes the last five years of accident data. However, within five years traffic may have varied, laws and regulations may have changed, and traffic facilities also may have altered. This study considers only three years of traffic information to provide a more accurate analysis.

A road file consists of five years of Utah Highway System route information. This data has not changed significantly during the last five years. The two files containing information on the vehicles and people involved in accidents were not used in this study.

6.3 Data Connection

The data necessary to calculate accident number or rate is contained in Table One (road information file) and Table Two (accident information file). The two tables must be considered together in order to connect the accident location to a certain route segment. The accident mile point must be compared with the begin-mile point and the end-mile point of the route segment to determine which segment the accident occurred in. The number of accidents occurring at each segment can be calculated separately.

6.4 Data Split

Before computing the range of the accident rate in each highway FC, it was necessary to split the records according to the FCs and to calculate the accident rate for each.

Structured Query Language (SQL) is a standard programming language for retrieving information from a database. It is used to build queries that connect, split, and retrieve data from different tables.
The following SQL statement retrieves the traffic accident table and the road table:

```
SELECT Count(*), [route_num], [begin_mp], [end_mp], [aadt], [functional_class],
count(*)*100000/([aadt]*365*3*([end_mp]-[begin_mp]))
FROM route_num
WHERE functional_class=1
GROUP BY [route_num], [begin_mp], [end_mp], [aadt], [functional_class]
ORDER BY [route_num], [begin_mp]
```

This command fulfills the following purposes:

1. Connects the two files according to the route number and mile point of the accidents.
2. Identifies the accidents occurring on the rural interstate highway (Highway FC 1).
3. Provides the total number of accidents for each road segment during the last three years.
4. Calculates the accident rate for each segment.

A similar SQL can be made for each highway FC.

7. DATA ANALYSIS

7.1 Preliminary Results

Through SQL statements, the rates for each highway segment were calculated and grouped according to FC. Figure 2 shows that the accident rate frequency for the rural interstate class has only one peak, but does not follow the normal distribution. The distribution of the accident rates is skewed to the right. The left tail (the lower accident rates) is tightly packed together while the right tail (the higher accident rates) is widely spread apart. This pattern occurred with all of the FCs.

Because this analysis is based on normalized shape distribution, it is necessary to transform the abnormal data into relatively normal data. Otherwise, the correct EV cannot be acquired. Possible approaches to working with the skewed distribution are discussed in Section 7.

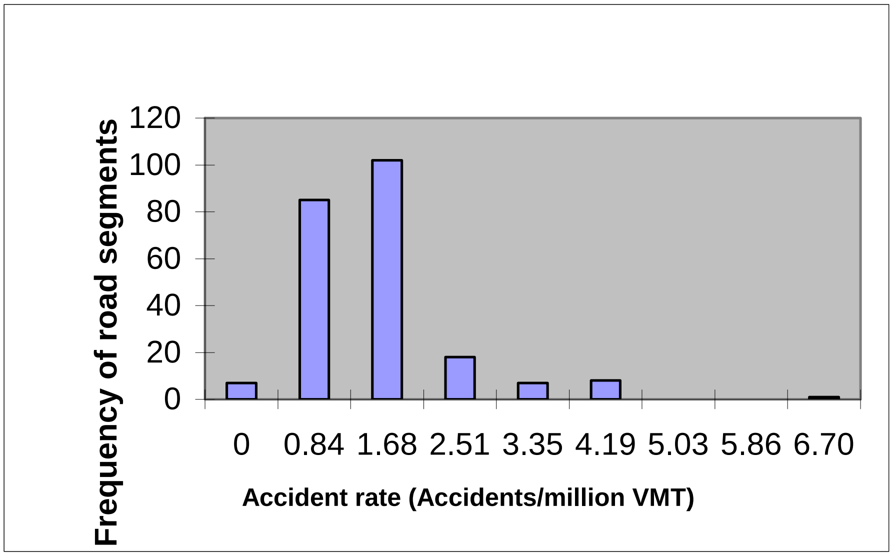

Figure 2: Accident Rate Distribution

7.2 Data Transformation

7.2.1 Data Trimming

Accident rates that are much higher than normal rates strongly impact average accident rate and standard deviation. These abnormally high rates are not consistent with other data and lead to higher expected accident rates. Data trimming removes the influence of abnormally high rates from the average, thus increasing accuracy in estimating the accident mean and standard deviation. When the original data set is trimmed, the mean value and the standard deviation are recalculated with the remaining data.

In this study, 10% of data was trimmed, 5% from each side. Figure 3 exhibits the balance of the data set's right and left tails after trimming.

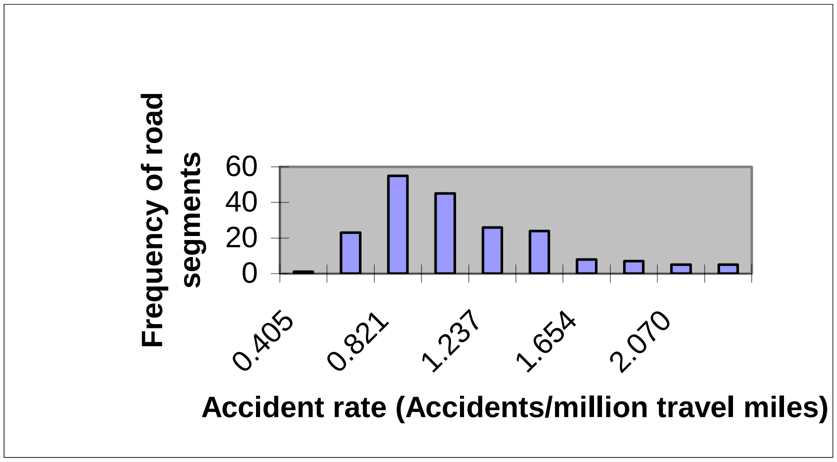

Figure 3: Accident Rate Distribution for Trimmed Data

7.2.2 Logarithm Transformation

Through mathematical transformations, this approach changes a data set that does not follow normal distribution into a data set with bell-shaped normal distribution. Logarithm Transformation is widely used in traffic research. It softens the impact of abnormally high statistics because it shrinks larger values more than smaller values.

Figure 4 shows a data set after logarithm transformation. The logarithm squeezes the right tail of the distribution and stretches the left tail, producing a greater degree of symmetry than existed in the skewed distribution. Squeezing high accident rates compensates for the right skew. Though the distribution is not perfectly bell-shaped, log transformation creates a more equal balance between the lower half than does the trimmed data set.

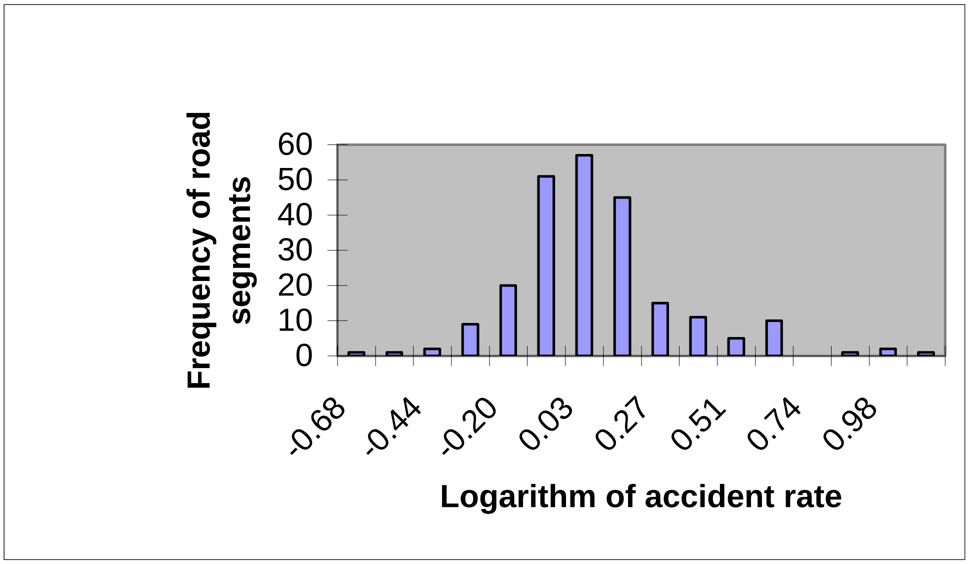

Figure 4: Accident Rate Distribution for Logarithm Data

7.3 Expected Values

Data transformations and deletion help distribution problems but do not provide a complete solution. In order to determine which works better for accident rate sets, these two transformations are compared with the original data. Therefore, three sets of data are compared and analyzed for each FC. These three data sets were applied to the EVA separately. The following table shows the EVs for the three sets of data.

Table 2: EVA Accident Rates for Utah Highway Functional Classes

Functional Class	Original Data			Log Data			Trim Data (5%)		
	Mean	Standard Deviation	Expected Value	Mean	Standard Deviation	Expected Value	Mean	Standard Deviation	Expected Value
Rural Interstate	1.25	1.26	**3.72**	1.03	1.79	**3.22**	1.08	0.54	**2.14**
Principal Arterial	2.07	3.14	**8.22**	1.51	2.28	**7.60**	1.72	1.41	**4.48**
Minor Arterial	2.52	4.52	**11.38**	1.72	2.37	**9.33**	1.82	1.74	**5.23**
Rural Major Collector	2.95	12.36	**27.18**	1.77	2.50	**10.66**	2.06	2.16	**6.29**
Rural Minor Collector	4.31	14.32	**32.38**	1.85	3.02	**16.14**	2.06	3.80	**9.51**
Rural Local	0.26	3.42	**6.96**	1.03	1.37	**1.91**	0.04	0.36	**0.75**
Urban Interstate	1.69	1.67	**4.96**	1.35	1.88	**4.65**	1.48	0.72	**2.89**
Urban Principal Arterial	2.03	1.77	**5.50**	1.57	2.09	**6.65**	1.78	0.89	**3.52**
Urban Principal Arterial, Other	6.34	8.66	**23.31**	3.89	2.72	**27.65**	5.17	3.83	**12.68**
Urban Minor Arterial	6.37	9.96	**25.89**	3.98	2.63	**26.48**	5.16	3.67	**12.35**
Urban Collector	7.06	16.78	**39.95**	3.31	3.06	**29.64**	4.55	4.81	**13.98**
Urban Local	2.40	10.86	**23.68**	1.43	2.42	**8.08**	0.84	2.20	**5.15**

Figure 5 shows the expected values for rural FCs according to original data, log data, and trimmed data. Figure 6 displays the expected values for urban FCs.

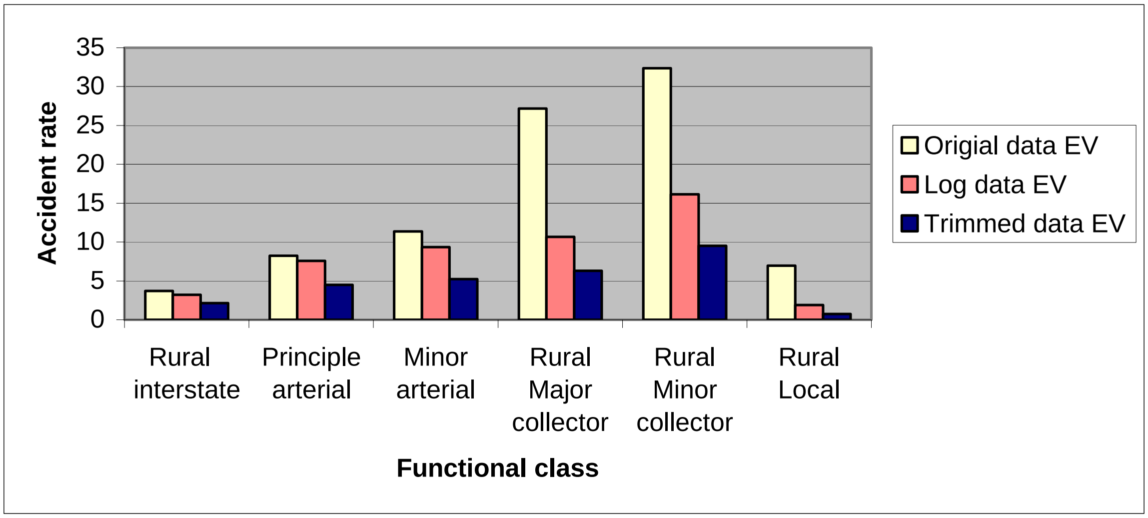

Figure 5: Expected Value of Accident Rate for Rural Functional Classes

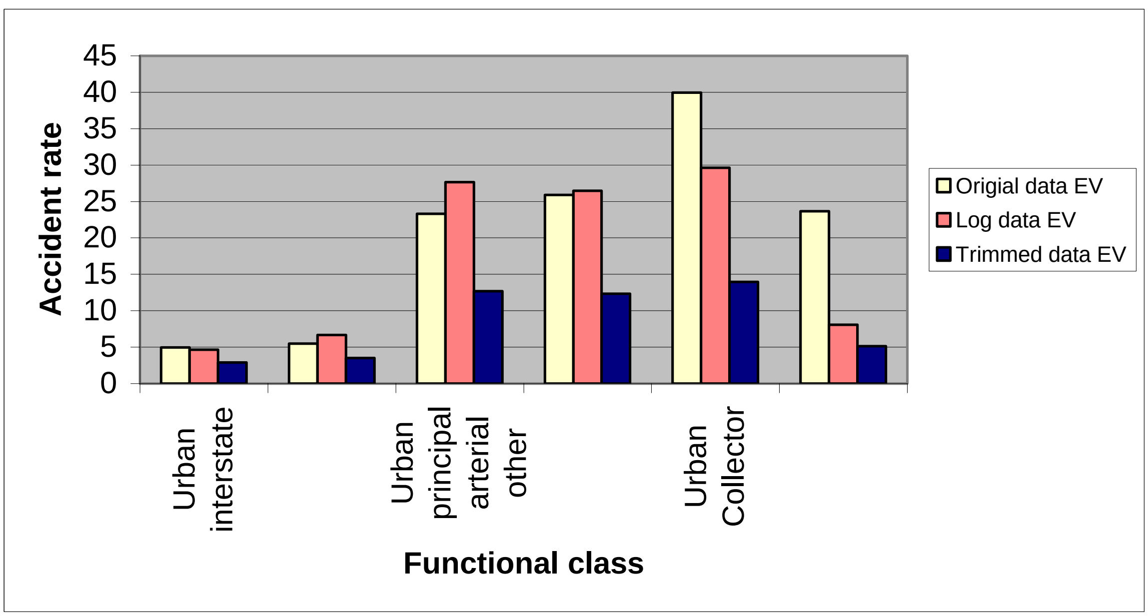

Figure 6: Expected Value of Accident Rate for Urban Functional Classes

7.4 Discussions

Table 2 shows that the collector and local classes have higher standard deviation values which cause the data to scatter. Standard deviation levels for the trimmed data sets in the FCs are much lower. Hence, the main reason for the high expected values of original data sets are the segments with abnormally high rates. However, before deleting the abnormal rates two questions should be asked. First, why should they be dropped? And, second, what percentage of the abnormal data should be removed?

The existing Utah highway functional classification system needs to be modified to include abnormal, or outlying data into its accurate FC. This process should include two steps. First, assume that the outlying data belongs to another FC. Second, check the abnormal data with accident data in other segments to determine if similar characteristics exist, such as traffic volume, location, length of segments, number of lanes, speed limit, and pavement type. If so, the outlying data may fit into a subclass of a certain FC. Because outlying data is no different than normal data it should not be trimmed, but rather placed in another FC.

Figures 5 and 6 show only slight differences among the values of log data sets and original data sets for rural interstate, rural principal arterial, rural minor arterial, urban interstate, urban principal arterial-freeways & expressways, and urban minor arterial. However for the rural major collector, the rural minor collector, rural local and urban principal arterial-other, urban collector, and urban local, the expected original data set values are much higher than those of log data sets. The EVs for original data sets for collector and local roads are high because the data are scattered. Distribution shapes for these classes are more similar to normal distribution after log transformation. Although log transformation does not completely follow normal distribution, it usually does not produce a greater symmetry.

7.5 Chi–square Test

Log transformation is a better approximation for the normal distribution than the other two data sets. Because EVA requires normal distribution of data sets, log transformation is the best way to transform original skew-distributed data. However, additional tests are required to check log transformation's capability to represent the original data sets.
The Chi-square test consists of the Null Hypothesis, or the theory that value sets fit a normal distribution, and the alternative hypothesis, or the theory that values do not fit a normal distribution.

H_0 = data set follows normal distribution ($X^2 <$ p–value).

H_1 = data set does not follow normal distribution ($X^2 >$ p–value).

The p-value measures the plausibility of the Null Hypothesis, which comes from the critical points of the Chi-square distribution table in this study.

Table 3 shows the Chi-square test applied to rural interstate and principal arterial FCs. It concludes the following:

1. None of the data sets follow normal distribution.
2. Log data sets have the smallest X^2 values, meaning that log data sets are more bell-shaped.

Therefore, log transformation can be adopted in future EVA.

Table 3: Chi-square Test Results

X^2	Original data	Log data	Trim data	P value
Interstate	179.19	38.95	66.15	25.19
Principal Arterial	264.31	95.51	178.87	26.76

8. CONCLUSIONS

1. According to the EVA results for accident rates, the Utah highway FC system is functional.
2. Table 4 uses EVs for accident rates to identify hazardous segments of each highway FC.
3. In the future, the EVs will vary according to changes in road geometrics, traffic volume, traffic speed, traffic control, and traffic density. Therefore, range tables will need to be modified accordingly. Figure 7 provides a method for modifying tables.

Table 4: Expected Values for FC Accident Rates

Functional class	Expected value
Rural interstate	3.22
Principle arterial	7.60
Minor arterial	9.33
Rural Major collector	10.66
Rural Minor collector	16.14
Rural Local	1.91
Urban interstate	4.65
Urban principal arterial	6.65
Urban principal arterial other	27.65
Urban Minor arterial	26.48
Urban Collector	29.64
Urban local	8.08

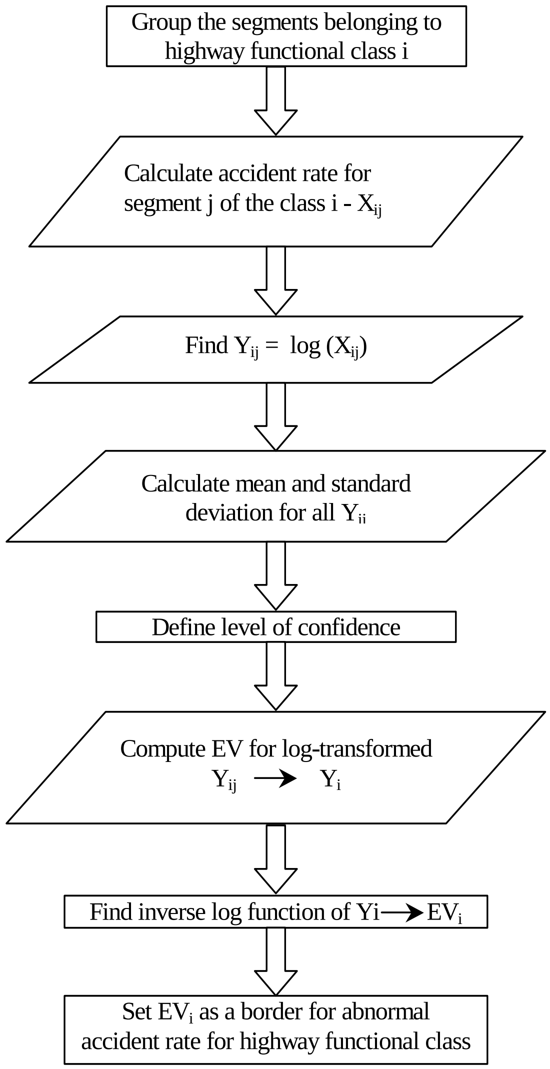

Figure 7: EVA process for Accident Rate

The following formula explains the process of finding EVA for rural interstate FC:

1. Connect the road table and the accident table.
SELECT [accident1999].[Milepoint], [road2001].[route_num], [road2001].[begin_mp], [road2001].[end_mp], [road2001].[functional_class], [road2001].[aadt], [road2001].[length_mp], [road2001].[county], [road2001].[city], [road2001].[district], [road2001].[maint_station], [road2001].[gov_level_control], [road2001].[num_lanes], [road2001].[spd_limit_mph], [road2001].[operating_spd_mph], [road2001].[peak_pct_truck], [road2001].[off_peak_pct_truck], [road2001].[urban_rural_desg], [road2001].[pavement_type], [road2001].[median_type], [road2001].[faid_system], [road2001].[faid_no], [road2001].[admin_class], [road2001].[route_signing], [road2001].[highway_type], [road2001].[nhs_system],

[road2001].[urban_location] FROM accident1999, road2001
WHERE ((([accident1999]![route_num])=[road2001]![route_num]) AND
(([accident1999]![milepoint])>=IIf([road2001]![begin_mp]=0,[road2001]![begin_mp],
[road2001]![begin_mp]+0.001) And
([accident1999]![milepoint])<=[road2001]![end_mp]));
Here, accident1999 is the accident table and road2001 is the road table. The SQL is used to combine them into one table, including the most important fields.
The results are saved as table **route_num**.

2. Select all of the rural interstate FCs from the road table and put them into another table.
SELECT * FROM road2001 WHERE ((([road2001].[functional_class])=1)) And [aadt]>10;
The results are saved as table **function1route**.

3. Calculate the accident number and accident rate for each segment belonging to rural interstate FC and summarize the count of the rows for each segment where functional_class=1.
SELECT Count(*), [route_num], [begin_mp], [end_mp], [aadt], [functional_class],
count(*)*1000000/([aadt]*365*3*([end_mp]-[begin_mp])) AS accident_rate
FROM route_num
WHERE functional_class=1 and aadt>10
GROUP BY [route_num], [begin_mp], [end_mp], [aadt], [functional_class]
ORDER BY [route_num], [begin_mp];
The result is saved as table **functionalclass1**.

5. Make a log transformation for each accident rate.

6. Calculate the mean of the log data and the standard deviation using data from table functionalclass1.

7. Compute EV for log data using Formula 1.

8. Find an inverse log value for each expected log value to find the border for the accident rate of each FC.

9. RECOMMENDATIONS

9.1 How to Transfer the Process to Oracle Database

Since UDOT will move the CARS data to the Oracle database, an automatic approach is recommended for EVA. The procedure follows.

1. Connect the road table and accident table and select the most important fields to form a new table.
2. Split the table into 12 sub-tables according to the FC.
3. Calculate the accident rate for each segment for a FC using one of the sub-tables.
4. Take a log transformation for the accident rate.
5. Compute the EV for log data.
6. Take an inverse log transformation to get the EV for accident rate.
7. Repeat the process 12 times for every FC.
8. Update the EV table.

The process in the Oracle database is similar to the process used in this project, however, a different form of SQL was used.

9.2 Automatic Process

All of these steps can be programmed into one single store procedure. Once the store procedure has been built in the Oracle database, users can easily calculate EV for different FCs.

REFERENCES

1. Bilal M. Ayyub, Richard H. McCuen, “Probability, Statistics, & Reliability for Engineers.”
2. H. Douglas Robertson, Joseph E. Hummer, Donna C. Nelson, “Manual of Transportation Engineering Studies.”
3. Applied Multivariate Data Analysis, volume I: Regression an Experimental Design
4. Robert W. Stokes and Madaniyo I. Mutabazi, TRR 1542, Rate-Quality Control Method of Identifying Hazardous Road Locations.
5. Ezra Hauer, Douglas W. Harwood, Forrest M. Council, HSRC, Michael S. Griffith, Estimating Safety by the Empirical Bayes Method: A Tutorial.
6. Higle, J.L., and Witkowski, J.M., Bayesian Identification of Hazardous Locations, *Transportation Research Record* 1185, TRB, 24-36, 1988.
7. T. Chira. Chavala and King K. Mak, Identification of Accident Factors on Highway Segments: A Method and Applications, TRR 1068.
8. MRRK PLASS and William D. Berg, Evaluation of Opportunity-Based Accident Rate Expressions, TRR.
9. King K. Mak, T.Chira-Chavala, and Barbara A. Hilger, Automated Analysis of High-Accident Locations, TRR 1068.
10. Northwestern University The traffic Institute traffic Engineering Analysis.
11. Identification of Abnormal Accident Patterns at Intersections, South Dakota Department of Transportation.
12. State Highway Crash Rate Tables, Oregon Department of Transportation.
13. Nicholas J. Garber, Lester A. Hoel, Traffic & Highway Engineering, Third Edition.
14. Fouad A. Gharaybeh, Identification of Accident-prone Locations in Greater Amman
15. Julia L. Higle, James M. Witkowski, Bayesian identification of hazardous locations